To: East Coast VLBI Workshop
From: John Gipson
Re: Inter-scan correlation: Threat or Menace
Date: October 13, 2005

Introduction

The VLBI observable is the delay between two stations. An important assumption that goes into the VLBI analysis is that all of the VLBI baseline delay observations are independent. This has the consequence that the uncertainty should be inversely proportional to the square root of the number of stations. The argument goes like this:

1. In least squares the uncertainty of an estimated parameter goes like:

$$\sigma_{est}^2 \approx \frac{NumEst}{NumObs}\sigma_{meas}^2$$

where σ_{meas}^2 is the measurement error and σ_{est}^2 is the error in the estimate.

2. The total number of observations in an experiment is roughly equal to the number of scans times the number of baselines. The number of baselines is Nstat*(Nstat-1)/2 (where Nstat is the number of stations).

3. The number of parameters grows linearly with the number of stations. Hence

$$\sigma_{Geo}^2 \approx \frac{1}{Nscan}\left(\frac{\sigma_{meas}^2}{Nstat}\right)$$

This does not seem to be case.

One possibility is that the observations on different baselines are not independent. If the observations were correlated, then our least squares estimates would under-estimate the formal errors. In general, there are two kinds of correlation that can arise:

1. Observations at different times may be correlated.
2. Observations at the same time may be correlated.

I will focus on the second.

My starting point is the hypothesis that there is station-dependent noise that leads to a change in the delay observable that is common to all baselines involving a given station on a given station.

Possible sources of such an effect are:

1. Atmosphere mis-modeling. Our model for the atmosphere is very simply. Although we do include azimuthal asymmetries in the form of gradients, the real atmosphere is more complicated. Errors in modeling the atmosphere will lead to station dependent delays.
2. Mismodeling the clocks.

3. Cable effects.
4. Any other effect that introduces a constant offset in delay at a given time.

There may also be station dependent noise that is not correlated across observations. This will just lead to an increased noise level. This is handled by re-weighting the data.

Station Dependent Delay in RDV48

My starting point in analysis is RDV48 (01 December 2005). This involved 16 stations total: Gilcreek, Hartrao, Medicina, Onsala, Tsukuba, Westford and the 10 VLBA stations. The motivation for using this session is:

1. Any station dependent delay effect will be more important the more stations are involved.
2. The RDVs generally involve some of the best stations. Hence these are good laboratories for studying weak effects.

I generated a standard solution where all of the re-weighting constants were set to zero. (The re-weighting constants were introduced to account for unmodeled noise. These constants are chosen so that the Chi-square of a VLBI solution=1.) One of my suspicions going in was that we are not doing the least-squares analysis correctly. In particular, that the observations involving a common station in scan will be correlated. If this is indeed true, this would lead to under-estimating the formal errors in a standard VLBI solution.

I had SOLVE write out the residuals to the solution. For each baseline delay observation SOLVE wrote out:

1. Source.
2. Time
3. Stations involved
4. Azimuth and elevation of the stations.
5. Delay Residual.
6. Uncertainty in the Delay measurement.
7. Rate Residual
8. Uncertainty in the Rate measurement.
9. Whether the observation is used in the solution.

In my analysis I only used observations that were used in the solution.

I went through the list of observations and grouped them into scans. This is not as easy as it sounds, because, observations in the same scan may have different time tags, and observations for different scans may be intermingled. This is particular true if there is sub-netting.

Table 1, below presents the residuals to the solution for a scan involving 1418+546. This is a 13 station scan.

1. The residual delay for a given baseline can be found by looking at the intersection of the rows and columns headed by station names.
    a. For example, the residual delay for FD-VLBA to HN-VLBA is 31 ps.

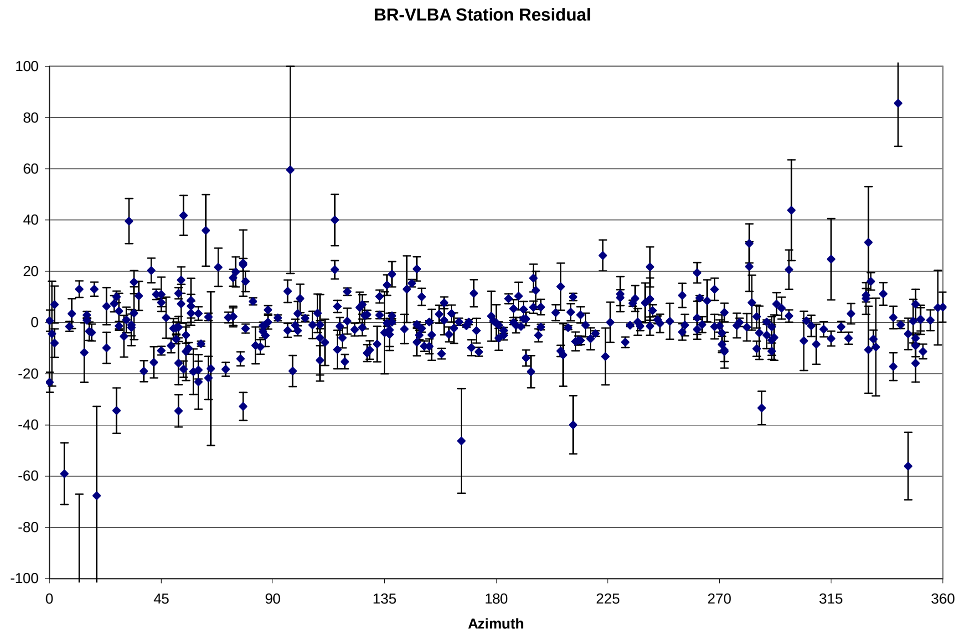
   b. The delay is anti-symmetric, so that the residual delay for HN-VLBA to FD-VLBA is -31 ps.
2. Cells containing a "-" indicate that either:
   a. The observation was not used in the solution.
   b. The observation was not correlated.
   c. The delay is not-meaningful, e.g. BR-VLBA to BR-VLBA.
3. The column labeld "GD" are the number of observations involving a particular station.
4. The column labeled AVG is the weighted average of the residual delay for all baselines to a station. I call this the **station-delay**. The station delay to HN-VLBA is -17.21 ps.
   a. The weights used in computing the average delay are the measurement errors, which are given in Table 2 for all baselines.
5. The column labeled SIG are the formal errors for the weighted delay residuals. These also use the measurement errors from Table 2.
6. The column labeled AVG/SIG is the average divided by the formal errors. This is a measurement of the significance of the station-delay.

A casual inspection of Table 1 shows that the residuals to a station tend to be correlated. For example, 10 of 12 the residuals to FB-VLBA are positive; 10 of the 12 residuals for NL-VLBA are negative.

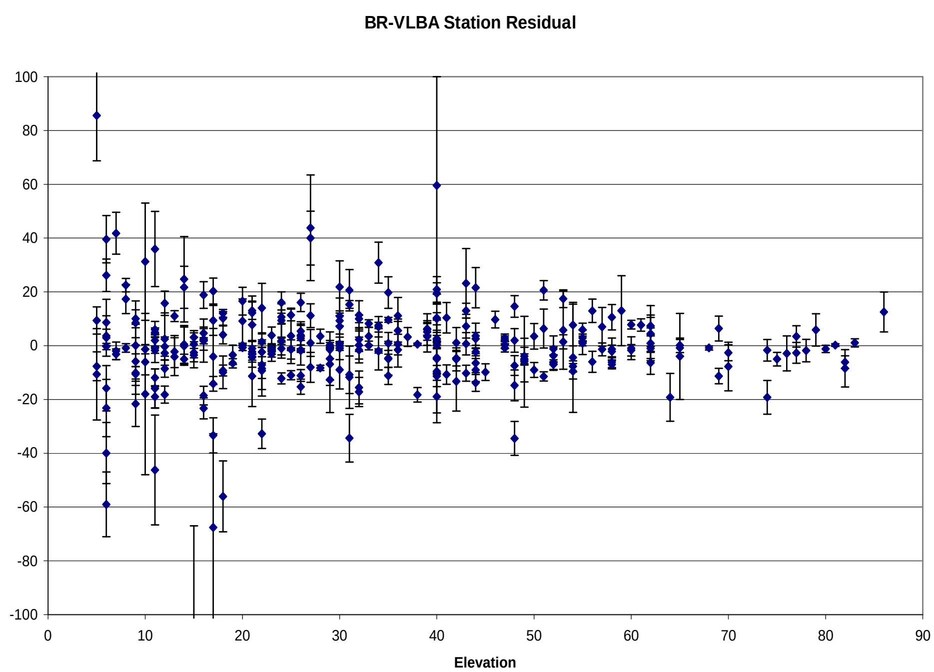
If the observations were truly independent, than we would expect that "on average" the residuals to a station would sum to 0. More precisely, we would expect that 2/3 of the time the sum of the residuals to a station would be within 1-sigma of 0. Only 3 of the 4 station delays are within 1 sigma of 0. In fact, 4 of the station-delays are at least 4-sigma away from 0. The probability of this happening by chance is nil.

Table 3 shows the residual delays corrected by subtracting (adding) the station delay determined from Table 1. After this correction, all but 1 of the AVG/SIGs is less than 1.

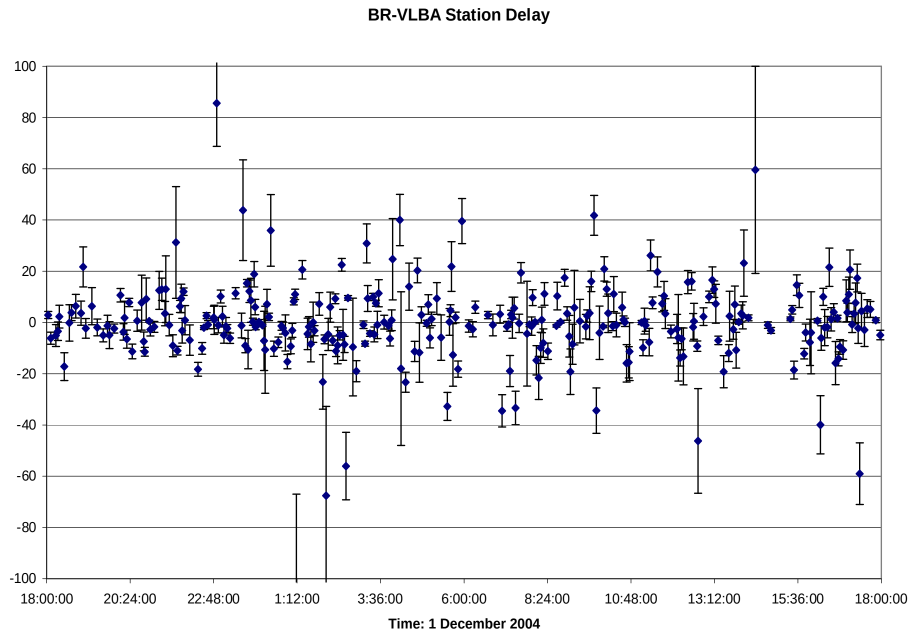
Although Table 1 concerns a particular scan, the results are more generally true. The next several figures display the station-residuals per-scan for BR-VLBA. The error bars are the computed formal errors using the uncertainties of the measurements.



The station delay does not seem to be depend on the azimuth.



There is some indication that the station-delay depends on the elevation.



The station-delay may be slightly correlated across time.

A natural question is what is the "average size" of the station-dependent delay. One measure of this is the RMS scatter in the station-dependent delay. Tables 4 and 5 calculate the weighted and unweighted RMS station delay for RDV48 and RDV50.

1. In general the weighted RMS is smaller. This is consistent with the above plots where the points with smaller error bars tend to be closer to 0.
2. The weighted RMS is on the order of 10 ps. This is significant.
3. The VLBA stations tend to have lower values than non-VLBA stations.

Table 4. RMS Variation of Station Dependent Delay: RDV48 December 1, 2004

	Station	# Scans	Wt RMS	RMS
1	BR-VLBA	292	8.6	16.3
2	FD-VLBA	301	8.6	20.5
3	GILCREEK	215	27.7	74.4
4	HARTRAO	152	31.8	56.8
5	HN-VLBA	248	13.5	27.0
6	KP-VLBA	294	6.2	10.2
7	LA-VLBA	301	7.2	11.7
8	MEDICINA	243	16.1	40.9
9	MK-VLBA	296	10.8	37.4
10	NL-VLBA	279	7.0	12.6
11	ONSALA60	162	18.2	43.3
12	OV-VLBA	295	7.6	14.0
13	PIETOWN	244	5.4	7.7
14	SC-VLBA	258	14.5	34.3
15	TSUKUB32	290	23.9	62.5
16	WESTFORD	221	17.7	26.7

Table 5. RMS Variation of Station Dependent Delay: RDV50 April 28, 2005				
	Station	# Scans	Wt RMS	RMS
1	ALGOPARK	173	5.6	23.5
2	BR-VLBA	83	5.8	12.1
3	FD-VLBA	229	6.8	12.9
4	GGAO7108	35	11.2	27.2
5	GILCREEK	146	11.8	40.4
6	HN-VLBA	189	6.7	31.2
7	KOKEE	171	11.3	23.1
8	KP-VLBA	225	7.6	15.4
9	LA-VLBA	220	8.7	18.5
10	MK-VLBA	190	8.0	21.0
11	NL-VLBA	221	6.0	21.6
12	OV-VLBA	214	7.0	26.3
13	PIETOWN	223	7.1	16.9
14	SC-VLBA	181	9.7	30.8
15	TIGOCONC	52	15.2	33.8
16	WESTFORD	157	8.4	31.1
17	WETTZELL	90	9.4	44.7

How to Solve This Problem & What are the Consequences

There are at (least) two ways of tackling this problem. One approach would be to just "noise-up" the sigmas that go-into the least squares estimates. For example, let $\sigma_{0,ij}$ be the naïve noise associated with an observation on baseline i-j. We could modify this by adding the expected station dependent noise for the two stations in quadrature:

$$\sigma^2_{new,ij}(new) = \sigma^2_{0,ij} + \sigma^2_i + \sigma^2_j$$

where σ^2_i is the station dependent noise for station *i*. This is actually the approach used in re-weighting the data to achieve a Chi-sq of one. The problem with this is that this ignores the correlations introduced by the station dependent noise. Although this will result in increasing the formal errors of the solution, the effect will not be large enough.

An alternate approach is to try estimate the station-dependent delay. It is worth noting that for an individual scan this delay is "clock" like. The station clocks can be estimated by assuming that the variation is piecewise linear. This imposes constraints on the behavior of the clocks. In contrast, the station-delays vary from scan to scan. The normal equations for the station dependent clocks are degenerate. In order to solve for these you need to impose constraints. One possibility is to loosely constrain them to 0, i.e., add in a pseudo-observations for each of the station-delay with uncertainty $\sigma_{S,i}$, where $\sigma_{S,i}$ is proportional to the RMS scatter of Tables 3 or 4. This allows you to invert the normal equations and has two additional consequences:

1. The formal errors of the geodetic quantities are increased because the number of parameters is increased.
2. This approach will automatically take into account the correlation between observations.

You can show under some special cases that if you use this approach, the formal errors for estimated geodetic quantities now goes like:

$$\sigma^2_{Geo} \approx \frac{1}{Nscan}\left(\frac{\sigma^2_{meas}}{Nstat} + \alpha\sigma^2_L\right)$$

Here α is just a proportionality constant. Note that as the σ^2_L goes to 0, we get the result at the start of the note. However, if this is non-zero this gives a lower limit on the uncertainty of estimated parameters.

Table 1. Residuals for Source 1418+546 @ 2004-12-01-18:06:59																				
#	Name	AZ	EL	BR-V	FD-V	HN-V	KP-V	LA-V	NL-V	OV-V	PIET	SC-V	WEST	MEDI	ONSA	TSUK	#GD	AVG	SIG	AVG/ SIG
1	BR-VLBA	322	82	-	-5	31	-20	-18	1	12	-21	8	36	-19	-11	-8	12	-6.14	2.37	-2.59
2	FD-VLBA	331	60	5	-	31	31	14	22	15	7	16	2	4	-39	-3	12	14.09	2.07	6.82
3	HN-VLBA	308	52	-31	-31	-	-6	-29	-6	-5	-22	-12	2	-12	-63	-31	12	-17.21	3	-5.73
4	KP-VLBA	338	65	20	-31	6	-	10	1	-45	-3	20	-11	-	-	46	10	-1.24	3.63	-0.34
5	LA-VLBA	327	66	18	-14	29	-10	-	2	-14	-3	-11	-22	-24	12	-24	12	-2.78	2.05	-1.36
6	NL-VLBA	310	63	-1	-22	6	-1	-2	-	-11	-14	20	-8	-17	-39	-17	12	-7.77	1.92	-4.05
7	OV-VLBA	341	71	-12	-15	5	45	14	11	-	14	-24	0	8	16	-36	12	-1.34	2.27	-0.59
8	PIETOWN	331	65	21	-7	22	3	3	14	-14	-	20	11	-21	16	-3	12	6.2	1.77	3.49
9	SC-VLBA	323	29	-8	-16	12	-20	11	-20	24	-20	-	6	4	64	23	12	-3.11	2.86	-1.09
10	WESTFORD	308	52	-36	-2	-2	11	22	8	0	-11	-6	-	34	-	32	11	0.24	4.34	0.06
11	MEDICINA	337	15	19	-4	12	-	24	17	-8	21	-4	-34	-	55	-28	11	2.81	4.31	0.65
12	ONSALA60	336	26	11	39	63	-	-12	39	-16	-16	-64	-	-55	-	-62	10	-13.56	11.95	-1.13
13	TSUKUB32	41	28	8	3	31	-46	24	17	36	3	-23	-32	28	62	-	12	13.88	3.45	4.02

Table 2. Uncertainties in Delay Measurements for Source 1418+546 @ 2004-12-01-18:06:59																	
#	Name	AZ	EL	BR-V	FD-V	HN-V	KP-V	LA-V	NL-V	OV-V	PIET	SC-V	WEST	MEDI	ONSA	TSUK	#GD
1	BR-VLBA	322	82	-	7	11	12	6	6	8	5	10	16	13	38	10	12
2	FD-VLBA	331	60	7	-	8	11	6	5	5	5	7	12	13	37	11	12
3	HN-VLBA	308	52	11	8	-	17	9	7	8	8	11	14	20	55	17	12
4	KP-VLBA	338	65	12	11	17	-	9	9	14	7	17	26	-	-	21	10
5	LA-VLBA	327	66	6	6	9	9	-	5	7	4	8	13	14	40	11	12
6	NL-VLBA	310	63	6	5	7	9	5	-	6	4	8	11	13	39	10	12
7	OV-VLBA	341	71	8	5	8	14	7	6	-	6	8	12	13	38	9	12
8	PIETOWN	331	65	5	5	8	7	4	4	6	-	7	11	12	34	9	12
9	SC-VLBA	323	29	10	7	11	17	8	8	8	7	-	16	16	47	15	12
10	WESTFORD	308	52	16	12	14	26	13	11	12	11	16	-	31	-	27	11
11	MEDICINA	337	15	13	13	20	-	14	13	13	12	16	31	-	36	10	11
12	ONSALA60	336	26	38	37	55	-	40	39	38	34	47	-	36	-	29	10
13	TSUKUB32	41	28	10	11	17	21	11	10	9	9	15	27	10	29	-	12

Table 3. Corrected Residuals for Source 1418+546 @ 2004-12-01-18:06:59

#	Name	AZ	EL	BR-V	FD-V	HN-V	KP-V	LA-V	NL-V	OV-V	PIET	SC-V	WEST	MEDI	ONSA	TSUK	#GD	AVG	SIG	AVG/ SIG
1	BR-VLBA	322	82	-	15	20	-15	-15	-1	17	-9	11	42	-10	-18	12	12	1.05	2.37	0.44
2	FD-VLBA	331	60	-15	-	0	16	-3	0	0	-1	-1	-12	-7	-67	-3	12	-2.3	2.07	-1.11
3	HN-VLBA	308	52	-20	0	-	10	-15	3	11	1	2	19	8	-59	0	12	0.67	3	0.22
4	KP-VLBA	338	65	15	-16	-10	-	8	-6	-45	4	18	-10	-	-	61	10	0.32	3.63	0.09
5	LA-VLBA	327	66	15	3	15	-8	-	-3	-13	6	-11	-19	-18	1	-7	12	0.48	2.05	0.24
6	NL-VLBA	310	63	1	0	-3	6	3	-	-5	0	25	0	-6	-45	5	12	1.35	1.92	0.7
7	OV-VLBA	341	71	-17	0	-11	45	13	5	-	22	-26	2	12	4	-21	12	1.15	2.27	0.51
8	PIETOWN	331	65	9	1	-1	-4	-6	0	-22	-	11	5	-24	-4	5	12	-1.75	1.77	-0.99
9	SC-VLBA	323	29	-11	1	-2	-18	11	-25	26	-11	-	9	10	54	40	12	0.72	2.86	0.25
10	WESTFORD	308	52	-42	12	-19	10	19	0	-2	-5	-9	-	37	-	46	11	-0.84	4.34	-0.19
11	MEDICINA	337	15	10	7	-8	-	18	6	-12	24	-10	-37	-	39	-17	11	1.77	4.31	0.41
12	ONSALA60	336	26	18	67	59	-	-1	45	-4	4	-54	-	-39	-	-35	10	2.17	11.95	0.18
13	TSUKUB32	41	28	-12	3	0	-61	7	-5	21	-5	-40	-46	17	35	-	12	-0.59	3.45	-0.17